university of groningen

University Medical Center Groningen

University of Groningen

Pitfalls of the Martini Model

Alessandri, Riccardo; Souza, Paulo C T; Thallmair, Sebastian; Melo, Manuel N; De Vries, Alex H; Marrink, Siewert J

Published in:
Journal of Chemical Theory and Computation

DOI:
10.26434/chemrxiv.8113172

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Final author's version (accepted by publisher, after peer review)

Publication date:
2019

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Alessandri, R., Souza, P. C. T., Thallmair, S., Melo, M. N., De Vries, A. H., & Marrink, S. J. (2019). Pitfalls of the Martini Model. *Journal of Chemical Theory and Computation*, *15*(10), 5448-5460. [acs.jctc.9b00473]. https://doi.org/10.26434/chemrxiv.8113172

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 05-05-2023

ChemRxiv™

doi.org/10.26434/chemrxiv.8113172.v1

Pitfalls of the Martini Model

Riccardo Alessandri, Paulo C. T. Souza, Sebastian Thallmair, Manuel N. Melo, Alex H. de Vries, Siewert-Jan Marrink

Submitted date: 15/05/2019 • Posted date: 16/05/2019

Licence: CC BY-NC-ND 4.0

Citation information: Alessandri, Riccardo; Souza, Paulo C. T.; Thallmair, Sebastian; Melo, Manuel N.; de Vries, Alex H.; Marrink, Siewert-Jan (2019): Pitfalls of the Martini Model. ChemRxiv. Preprint.

The computational and conceptual simplifications realized by coarse-grain (CG) models make them an ubiquitous tool in the current computational modeling landscape. Building block based CG models, such as the Martini model, possess the key advantage of allowing for a broad range of applications without the need to reparametrize the force field each time. However, there are certain inherent limitations to this approach, which we investigate in detail in this work. We first study the consequences of the absence of specific cross Lennard-Jones parameters between different particle sizes. We show that this lack may lead to artificially high free energy barriers in dimerization profiles. We then look at the effect of deviating too far from the standard bonded parameters, both in terms of solute partitioning behavior and solvent properties. Moreover, we show that too weak bonded force constants entail the risk of artificially inducing clustering, which has to be taken into account when designing elastic network models for proteins. These results have implications for the current use of the Martini CG model and provide clear directions for the reparametrization of the Martini model. Moreover, our findings are generally relevant for the parametrization of any other building block based force field.

File list (2)

pitfalls-achemso.pdf (892.55 KiB) view on ChemRxiv • download file

pitfalls-achemso-suppinfo.pdf (9.24 MiB) view on ChemRxiv • download file

Pitfalls of the Martini Model

Riccardo Alessandri,[†,‡] Paulo C. T. Souza,[†,‡] Sebastian Thallmair,[†] Manuel N. Melo,[¶] Alex H. de Vries,[†] and Siewert J. Marrink[*,†]

[†]*Groningen Biomolecular Sciences and Biotechnology Institute and Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands*

[‡]*These two authors contributed equally.*

[¶]*Instituto de Tecnologia Química Biológica António Xavier, Universidade Nova de Lisboa, Av. da República, 2780-157 Oeiras, Portugal*

E-mail: s.j.marrink@rug.nl

Abstract

The computational and conceptual simplifications realized by coarse-grain (CG) models make them an ubiquitous tool in the current computational modeling landscape. Building block based CG models, such as the Martini model, possess the key advantage of allowing for a broad range of applications without the need to reparametrize the force field each time. However, there are certain inherent limitations to this approach, which we investigate in detail in this work. We first study the consequences of the absence of specific cross Lennard-Jones parameters between different particle sizes. We show that this lack may lead to artificially high free energy barriers in dimerization profiles. We then look at the effect of deviating too far from the standard bonded parameters, both in terms of solute partitioning behavior and solvent properties. Moreover, we show that too weak bonded force constants entail the risk of artificially inducing clustering, which has to be taken into account when designing elastic network models for proteins. These

results have implications for the current use of the Martini CG model and provide clear directions for the reparametrization of the Martini model. Moreover, our findings are generally relevant for the parametrization of any other building block based force field.

1 Introduction

Coarse-grain (CG) models play an increasingly important role in computational science, and are nowadays a tool as important as atomically detailed models.[1–6] By grouping atoms into effective interaction sites, often called *beads*, CG models focus on essential features, while averaging over less vital details. This provides significant computational and conceptual advantages compared to more detailed models, allowing to probe the temporal and spatial evolution of systems on the mesoscale.

Among the philosophies of CG modeling, we find both systematic (also known as hierarchical) and building block approaches.[2,3,5] CG models developed on the basis of the former, purely "bottom-up" principle focus on the accurate reproduction of the underlying atomistic structural details at a particular state point for a specific system, but require reparametrization whenever any condition changes. This translates into a more time-consuming parametrization procedure. Moreover, potential forms required are often complex, which can result in lower performance and thus less sampling. On the other side, building block approaches usually rely more heavily on a "top-down" approach, where macroscopic properties (*e.g.*, thermodynamic data) are used as the main target of their parametrization. "Top-down" CG models are often cheaper—due to simpler potential forms and only partial parametrization required—and transferable, as the parametrization of the building blocks allows to re-use them as part of similar moieties in different molecules. However, the structural accuracy of top-down models is limited as the representation of the atomistic detail is suboptimal. The line that separates these two methodological philosophies is, however, thin. Many successful force fields have been developed combining top-down and bottom-up approaches.[3,5]

One important example of the building block philosophy applied to CG modeling is the Martini force field.[7,8] Designed as a model for simulations of lipids and surfactants,[9] this force field has become the most widely-used CG model for simulations of biomolecules,[8] and it is increasingly popular in soft material science.[10–15] The Martini model mainly relies on a four-to-one mapping scheme, where on average four non-hydrogen atoms are mapped into a CG regular (R) bead. Finer mappings of up to two-to-one non-hydrogen atoms-to-CG-site are employed when the symmetry of the molecules requires it or for ring-like structures. In the latter cases, small (S) or tiny (T) CG beads are employed.[7,16] There exist four main types of particles: polar (P), non-polar (N), apolar (C) and charged (Q). These types are in turn divided in subtypes based on their hydrogen-bonding capabilities (with a letter denoting: d = donor, a = acceptor, da = both, 0 = none) or their degree of polarity (with a number from 1 = low polarity to 5 = high polarity). This gives a total of 18 particle types: the *Martini building blocks*. The "flavor" of each building block is determined by the non-bonded interactions, which are described by a Lennard-Jones (LJ) 12-6 potential:

$$V^{\mathrm{LJ}}(r_{ij}) = 4\varepsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] \tag{1}$$

The LJ σ parameter, determining the effective size of the beads, is 0.47 nm for regular interactions. For the smaller sizes, it is reduced to 0.43 nm and 0.32 nm in the case of S-S and T-T interactions, respectively. The LJ well-depth ε parameters, determining the strength of the interactions between bead pairs, can vary from 5.6 kJ mol^{-1} to 2.0 kJ mol^{-1}. These values are scaled down by 75% in the case of S-S or S-T interactions. Together, these LJ parameters determine how the building blocks interact with each other, giving rise to the Martini interaction matrix.[7] Non-bonded interactions were parametrized based on thermodynamic data describing the different affinities of chemical groups towards different solvent phases, namely, free energies of transfer between water and a number of organic phases ("top-down").[7] Bonded interactions, described by a standard set of potential energy functions

common in classical force fields, are parametrized from the underlying atomistic geometry, usually comparing to atomistic simulations ("bottom-up") or experimental data. This seemingly simple approach, a thorough parametrization of the hydrophilicity/hydrophobicity of the building blocks of the model, resulted in a wide range of successful applications in the modeling of (bio)molecular processes.[8]

The parametrization of any (building block based) (CG) force field is performed under a set of specific conditions. Parametrizations are carried out on a necessarily limited set of systems described by a number of standard parameters such as the range of LJ parameters and bond lengths, and assuming a number of simulation settings which specify how the simulations are carried out, such as the treatment of interactions between particles and temperature and pressure coupling schemes. In the case of the Martini force field, the parametrization was mainly carried out using isolated regular beads or linear molecules composed of such beads, and a number of standard parameters for the models, such as bond lengths, and angles.[7] Moreover, specific settings were employed for the treatment of the interactions between particles, such as cutoff settings. The latter settings will not be discussed in the present work and the interested reader is referred to Ref. 17 for a recent work discussing these settings in Martini. Overall, the conditions employed during the parametrization allow for more or less freedom, but there are always boundaries.

Here, we investigate what can happen when pushing the limits of the parametrization of a building block based CG model, with focus on the Martini force field. Given its very wide use, its (necessarily) more modest initial boundaries of parametrization have been pushed to their limits. In particular, section 3.1 discusses problems arising from the lack of size-dependent Lennard-Jones interaction parameters. We then explore how going too far from the original bonded parameters affects the behavior of the force field, both in terms of solute (section 3.2) and solvent phases (section 3.3). In section 3.4, we then demonstrate how (effective) bond length distributions can be affected due to the application of an elastic network, with consequences for the behavior of the model. Finally, section 4 concludes

discussing the implications for the use of the current version of the Martini force field and directions for reparametrizations.

2 Methods

All-Atom and Coarse-Grained Models. The benzene all-atom (AA) models used in Figure 1 are standard GROMOS (53A6)[18] (retrieved from the ATB server[19]) and OPLS[20] models. Standard Martini 2.2 models[7,16,21] (available on the Martini portal http://cgmartini.nl) were used for the solvents considered in Figure 4.

Simulation Settings. A unique set of GROMACS atomistic run parameters was used for the AA simulations. The Verlet neighbor search algorithm was employed to update the neighbor list, and a 1.4 nm cutoff for LJ and for Coulomb (reaction-field) interactions was employed. The Nosé-Hoover[22,23] thermostat (coupling parameter of 1.0 ps) and the Parrinello–Rahman barostat[24] (coupling parameter of 5.0 ps) were employed to maintain temperature (298.15 K) and pressure (1 bar), respectively. A stochastic integrator was employed for both AA and CG simulations. Settings for the CG simulations follow the "new" Martini set of run parameters.[17] Specifically, the Verlet neighbor search algorithm is used to update the neighbor list, with a straight cutoff of 1.1 nm. The velocity-rescaling thermostat[25] (coupling parameter of 1.0 ps) and the Parrinello–Rahman barostat[24] (coupling parameter of 12.0 ps) were employed to maintain temperature (298.15 K) and pressure (1 bar), respectively. CG simulation setting files are available on the Martini portal http://cgmartini.nl. GROMACS[26] 2016.x was employed to run the simulations.

Potential of Mean Force Calculations. The Potential of Mean Force (PMF) profiles were obtained from umbrella sampling simulations.[27] The two solute molecules (either an atomistic or a Martini benzene model or single Martini beads) were placed in a box (of at least $5 \times 5 \times 5$ nm^3) and solvated in water (using the SPC[28] and the TIP3P[29] water models in the GROMOS and OPLS case, respectively, and the standard Martini water model at the

CG level[7]). Umbrella windows were spaced 0.1 nm apart along the reaction coordinate, this being the distance between the centres of mass of the solute molecules. In each window, the distance was restrained by applying a harmonic potential with a force constant of 1500 kJ mol^{-1} nm^{-2}. Each window was equilibrated for 3 ns (1 ns) and then simulated for 500 ns (150 ns) in the CG (AA) case. A stochastic integrator was employed, while other settings were the same as the general settings described above. The free energy profiles were calculated using the weighted histogram analysis method (WHAM)[30] as implemented in the GROMACS tool `gmx wham`.

Free Energies of Transfer Calculations. Thermodynamic integration (TI) was used to compute free energies of solvation $\Delta G_{S\to\emptyset}$ in a solvent S. The solute was solvated in a pre-equilibrated solvent box of size of at least $5 \times 5 \times 5$ nm^3. A series of 11 simulations with equally spaced λ points going from 0 to 1 were performed. A stochastic integrator was employed, and simulations were equilibrated for 2 ns and each λ point was run for 10 ns. A soft-core potential (with α of 0.5 and power set to 1) was employed to avoid the singularity in the potential when interactions were switched off.[31] The free energies and corresponding errors were finally computed using the Multistate Bennett Acceptance Ratio (MBAR).[32] The free energy associated with transferring a solute from a solvent S_1 to a solvent S_2 ($\Delta G_{S_1\to S_2}$) is then computed as the difference $\Delta G_{S_1\to\emptyset} - \Delta G_{S_2\to\emptyset}$. In the *repulsive* TI calculations (*e.g.*, Figure 3) the solvent-solute attractive interactions are switched off, that is, the solute-solvent Lennard-Jones dispersion constant C_6 is set to zero. The free energies obtained for placing such a purely repulsive LJ particle in a solvent phase captures the free energy cost of creating a cavity in that solvent.

Enthalpies of Vaporization Calculations. The enthalpy of vaporization (ΔH_{vap}) has been computed according to:

$$\Delta H_{\text{vap}} \approx U_{\text{gas}} - U_{\text{liq}} + RT \tag{2}$$

where U_{gas} and U_{liq} are the total energies (per mole) of the gas and liquid phase, respectively. The gas phase is approximated as one molecule in a large ($7 \times 7 \times 7$ nm^3) empty simulation box, and the liquid phase as an equilibrated box of dimensions of about $5 \times 5 \times 5$ nm^3. Gas (liquid) phase simulations were performed in the NVT (NPT) ensemble at 298 K (and 1 bar).

1:1 Mixture and Icosahedron System Setup. The 1:1 mixture of dodecane and dodeca-1,3,5,7-tetraene was set up using the `gmx insert-molecules` tool of GROMACS. 350 molecules each were added to a rectangular box of $3.5 \times 3.5 \times 20$ nm^3. The starting configuration of the eight icosahedrons in a cubic box ($8.5 \times 8.5 \times 8.5$ nm^3) was generated using the `gmx insert-molecules` tool to add the eight icosahedrons and the `gmx solvate tool` to solvate the system with 4634 CG water beads. Both systems were energy minimized (steepest descent, 500 steps), equilibrated for 2 ns (time step of 20 fs), and simulated for 500 ns (time step of 20 fs). A leap-frog integrator was employed.

Polyleucine System Setup. To embed the nine polyleucine peptides in the POPC bilayer, the program `insane.py` was employed.[33] The peptides were placed on a cubic grid at a distance of 5 nm resulting in a bilayer patch of 15×15 nm^2 and an overall box size of $15 \times 15 \times 10$ nm^3. The POPC bilayer consisting of 567 lipids was solvated with 14211 CG water beads and 0.13 M NaCl was added after neutralizing the system. The system was then energy-minimized (steepest descent, 500 steps), equilibrated for 500 ps (time step of 10 fs), and simulated for 10 μs (time step of 20 fs). A leap-frog integrator was employed, and reaction-field was used for Coulomb interaction (cutoff 1.1 nm)—as this system is charged, following the "new-rf" set of Martini run parameters.[17] Two different polyleucine (LYS_2-$LEU_2$6-LYS_2) protein models were used: a standard Martini 2.2 model without any elastic network and a Martini 2.2 model with elastic network of GROMACS bond type 1. Bond type 1 in GROMACS means that the non-bonded interactions between the connected beads are excluded. In both protein models, identical regular bonds, angles, and dihedral angles are applied; their only difference is the elastic network. This allows us to exclude any changes

due to different bonded parameters. However, this is a setting for the elastic network which is not commonly used when simulating proteins with Martini and an elastic network. There are two elastic network options commonly used: Martini 2.2 with elastic network of GROMACS bond type 6, which does not exclude the non-bonded interactions, or the ElNeDyn[34] model, which uses GROMACS type 1 bonds in the elastic network but entails in addition different definitions of bonded interactions.

3 Results and discussion

3.1 Differences in Bead Sizes: the Desolvation Problem

Different Bead Sizes Can Lead to Artificial Free Energy Barriers. Mixing different particle sizes *without* introducing also mixed resolution LJ parameters can lead to artificial free energy barriers. This is the case in Martini when interactions between small or tiny and regular beads take place. As described above, the LJ σ parameter in the case of S-S interactions is reduced to 0.43 nm, from the 0.47 nm used for R-R interactions. At the same time, the LJ ε between two S-particles, $\varepsilon_{\text{S-S}}$, is also reduced by a quarter as compared to the interaction between two regular Martini particles, *i.e.*, $\varepsilon_{\text{S-S}} = 0.75\ \varepsilon_{\text{R-R}}$. In the case of T-beads an even smaller $\sigma_{\text{T-T}}$ of 0.32 nm is used, while no $\varepsilon_{\text{T-T}}$ scaling is applied. However, for simplicity, the LJ ε and σ for R-S (R-T) interactions are kept the same as the ones for R-R interactions. Therefore, S-beads (T-beads) are seen by R-beads as regular particles, while they interact with other S-particles (T-particles) with the reduced σ values. We find that this leads to the formation of artificial free energy barriers. This can be observed in, for example, potentials of mean force (PMFs) of dimerization of molecules described by S- or T-beads solvated in a R-bead solvent. Such a case is shown in Figure 1, where PMFs of dimerization of two Martini three-particle ring molecules in water are shown for the three different bead sizes available in the force field. The PMFs have been computed by Umbrella Sampling, as described in the Methods section.

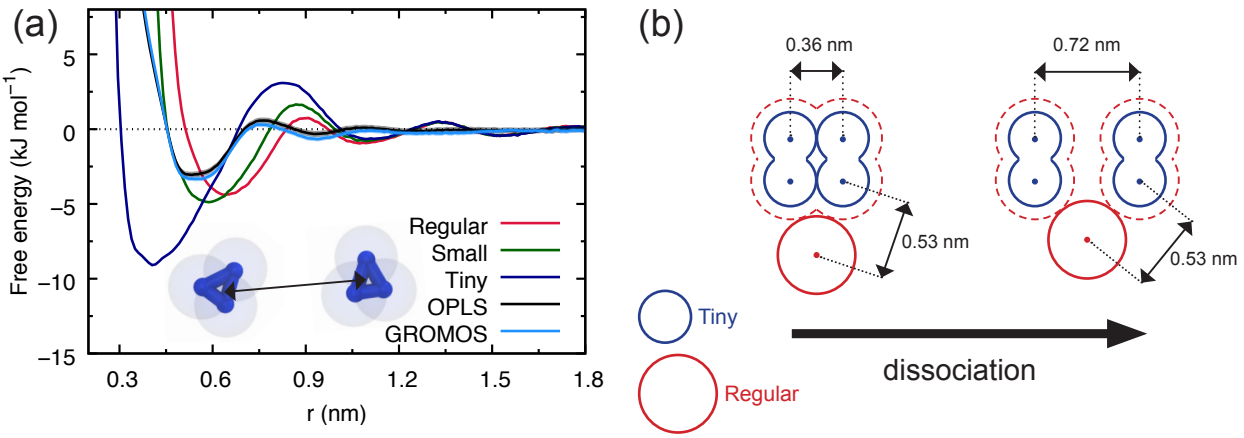


Figure 1: Effect of lack of size-dependent Lennard-Jones parameters between particles of different sizes on the dimerization. (a) Potentials of mean force for the dimerization of three-particle Martini ring molecules (see inset) described by regular (red), small (green) or tiny (blue) beads, and for atomistic GROMOS (cyan) and OPLS (black) benzene models in water. (b) Schematic of the T-solute in R-solvent dissociation (side view): because the solute is seen by the solvent molecules as described by regular beads, a solvent molecule cannot insert between the two solute molecules until the distance between the positions of the beads is at least twice the diameter of a regular bead.

The LJ ε value for the self-interaction of the solute molecules is kept constant in the three cases (we chose the scaled down intermediate level IV, *i.e.*, 2.625 kJ/mol), so as to exclude its effect. Bond lengths between the ring particles are also constrained in the three cases, and are set to 0.27 nm, the bond distance used in the standard Martini benzene model.[7] Atomistic PMFs, computed employing the GROMOS (53A6)[18] and OPLS[20] force fields, are also shown in Figure 1a for the dimerization of benzene molecules, which the S-ring model may be taken to represent. It is very evident from the plot that, going from the R-ring to its T-version, an energy barrier arises at around 0.8 nm, while no such barrier is present in the atomistic PMFs. The barrier increases as the difference in size between the solvent and solute beads increases.

Lack of Size-Dependent Cross Interactions. We rationalize the appearance and increase of the barrier by looking at a simple picture representing the system, a schematic of which is reported in Figure 1b. Note that, despite referring to the T-systems solvated in water, the following description applies generally to all (T-, S-, R- and atomistic) systems, as

a barrier is present in all cases. However, using the R-R LJ parameters for the R-S and R-T LJ cross interactions artificially increases the height of the barrier in the case of the S- and T-systems. When the two T-molecules are in close contact—which happens at about 0.36 nm, *i.e.*, the diameter of a T-bead—the interaction is favorable, and the free energy is at its absolute minimum on the profile shown in Figure 1a. As the two T-molecules are pulled apart, a cavity starts to form between them. This cavity cannot be filled by any solvent molecule until the distance between the two solute molecules becomes equal or larger than the diameter of interaction dictated by the solute-solvent LJ σ parameter (*i.e.*, the σ_{R-T} parameter in this case). This translates into an energy barrier, which is generally observed in conjunction with dimerization.[35] However, for a solvent R-bead to make its way between the two T-systems the distance between the positions of the beads must be not only 0.89 nm (that is, the diameter of a T-bead, 0.36 nm, plus the diameter of a R-bead, 0.53 nm), but instead 1.06 nm (that is, twice the diameter of a R-bead), as the T-systems are seen as composed by regular beads by the solvent particles. This translates into the formation of a cavity which is larger than what it would be if the σ_{R-T} was tailored for R-T interactions (*e.g.*, one could take the arithmetic or geometric average between the σ_{R-R} and σ_{T-T}), and which thus has an associated increased cavity cost. This leads to the artificially higher free energy barrier observed for the S- and T- systems, which is higher the larger the mismatch between the solute-solute and solute-solvent σ parameters is.

Generality and Consequences. The formation of such an artificial energy barrier has been shown for the case of a prototypical CG ring molecule. The effect is most obvious in Martini for ring systems that use S- and T-beads. However, the effect is not limited to ring geometries. It also plays a role in the simplest case, *i.e.*, the dimerization of single beads. The PMFs obtained in this case are shown in Figure S1a in the Supporting Information, and demonstrate again the appearance of an energy barrier as the difference in size between solute and solvent increases—only smaller, as fewer particles are involved. While free energy barriers are usually observed in conjunction with dimerization,[35] the increase of such

free energy barriers in Martini is a direct consequence of the lack of size-dependent cross interactions. It is thus a consequence of the design of the model itself one should be aware of. In the T-case, the situation is evidently problematic, and this effect is very noticeable in the stacking and base-pair PMFs of the Martini DNA bases.[16] It should be noted that in case of small S-bead solutes (up to a few particles), the effect is relatively mild (compare the green curve to the atomistic case in Figure 1a). However, as the number of particles in the ring structure increases, the effect increases, as can be seen from the PMF of the polycyclic aromatic molecule pyrene (Figure S1b in the Supporting Information).

3.2 Solutes with Short Bond Lengths: Effects on Oil/Water Partitioning

We now look at the effects of bonded parameters on the behavior of the Martini CG force field. In particular, we investigate the robustness of the Martini model upon changes in the bond lengths which connect the various building blocks. To this end, we systematically study the effect that varying the bond lengths has on the reproduction of the main experimental parametrization target of the Martini model, *i.e.*, the partitioning behavior of molecules. This is done by comparing changes in experimental and computed free energies of transfer ($\Delta G_{\text{transfer}}$) upon changes in bond lengths. It is useful to define the change in the free energy of transfer ($\Delta\Delta G_{\text{transfer}}$) as follows:

$$\Delta\Delta G_{\text{transfer}} = \Delta\Delta G_{\text{solute-solvent}} + \Delta\Delta G_{\text{solvent-solvent}} \tag{3}$$

that is, we can divide the change in the free energy of transfer in contributions due to solute-solvent, and solvent-solvent interactions. In this work, we discuss uncharged systems, so the interactions involved in Eq. 3 are controlled by the σ and ε LJ parameters (Eq. 1) associated with the solute and solvent particles. Bond lengths, along with the LJ σ, determine *the density of interaction sites* that will be found in the simulation. In turn, the density of

interaction sites affects the strength of the interactions between molecules, and therefore their behavior (*i.e.*, thermodynamic properties). The LJ ε parameters between the building blocks of the Martini model were parametrized mostly based on single R-beads or molecules composed of linear R-bead chains employing a standard bond length of 0.47 nm. In this section, we vary the bond lengths of the *solute* molecule, while using Martini solvents either consisting of single R-beads (like water) or described by models composed of linear R-bead chains with standard bond lengths of 0.47 nm (like hexadecane). Thus, we first look at the impact of shortening bond lengths on the $\Delta\Delta G_{\mathrm{solute-solvent}}$ of Eq. 3, while using standard Martini solvents and thus well-calibrated solvent-solvent interactions.

Experimental Behavior Corresponding to Shortening Bond Lengths. Before describing the results, it is instructive to consider what changing a bond length in a (Martini) CG model means in terms of the actual molecules, and what behavior(s) should be thus captured by the model. Shorter CG bond lengths arise when the number of atoms mapped with the same number of beads is lower (*e.g.*, when a two-bead model to describe octane is adapted to heptane), or when the molecule is branched (*e.g.*, going from octane to tetramethylbutane). We focused on studying the partitioning behavior of molecules upon removal of aliphatic carbon atoms, as this corresponds to shortening bond lengths in a (Martini) CG model. We gathered a large set of partitioning data[36] from where to extract experimental trends. The data are plotted in Figure 2a, and show how the hexadecane→water free energy of transfer ($\Delta \mathrm{G}_{\mathrm{HD}\rightarrow\mathrm{W}}$), for the same chemical functional group, changes upon removal of aliphatic carbons (for an example, see Figure 2b). This particular free energy of transfer has been chosen because it comprises the two prototypical extremes of hydrophobicity and hydrophilicity, and because there are numerous experimental data points available. It is evident that the hydrophilicity of molecules increases upon reduction of their size, which is the case when removing atoms. This is to be expected, given the higher cost of creating a cavity in water as compared to hexadecane[35] which translates into a higher hydrophilicity of the molecule upon size reduction, as the free energy gain in creating a smaller cavity in

water outweighs the one in hexadecane.

Overall, the main effects of reducing the size of a molecule can be summarized as shown in Figure 2b: smaller molecules interact less with the environment, and possess reduced solvent accessible surface area; due to their smaller size, their solvation comes with a lower ΔG_{cavity} in any solvent; due to the high cost of creating a cavity in water, a greater discount is obtained on the ΔG_{cavity} in water upon size reduction, which makes smaller molecules more hydrophilic. A quantitative empirical observation can also be extracted: hydrophobic molecules get from 3.0 to 3.5 kJ mol^{-1} more hydrophilic for each aliphatic carbon atom removed, while hydrophilic molecules get a somewhat smaller free energy gain (2-2.5 kJ mol^{-1}). That the proximity of an aliphatic carbon atom to a polar group alters its hydrophobicity is in line with the proximity-based correcting factors often applied within partition coefficients (logPs) prediction schemes.[37]

Effect of Bond Lengths on the Partitioning of Martini Molecules. We now turn to the CG model, to investigate whether it succeeds in capturing the experimental partitioning behavior of molecules upon size reduction. Computed free energies of transfer (via thermodynamic integration as described in the Methods section) are plotted in Figure 2c, in a similar way to what was done in Figure 2a. In this case, the removal of 1, 2, and 4 aliphatic carbon atoms corresponds to model bond lengths of 0.4, 0.3 and 0.2 nm, respectively, and the free energies on the horizontal axis are the ones for all possible Martini two-bead "standard" molecules (*i.e.*, 4 atoms-to-1 CG site mapped molecules with a bond length of 0.5 nm—note that the standard Martini bond length is 0.47 nm, but there is no significant difference between using 0.47 or 0.5 nm). Note that the removal of 4 aliphatic carbon atoms is an extreme case: it is not realistic to remove 4 carbon atoms and stick to a two-bead model. The Martini approach would dictate that one bead gets removed at that point. However, it is instructive to push this far to extract trends in the behavior of the force field.

Overall, with respect to the experimental behavior, all the molecules with shorter bond lengths become less hydrophilic than what they should (compare to Figure 2a). However,

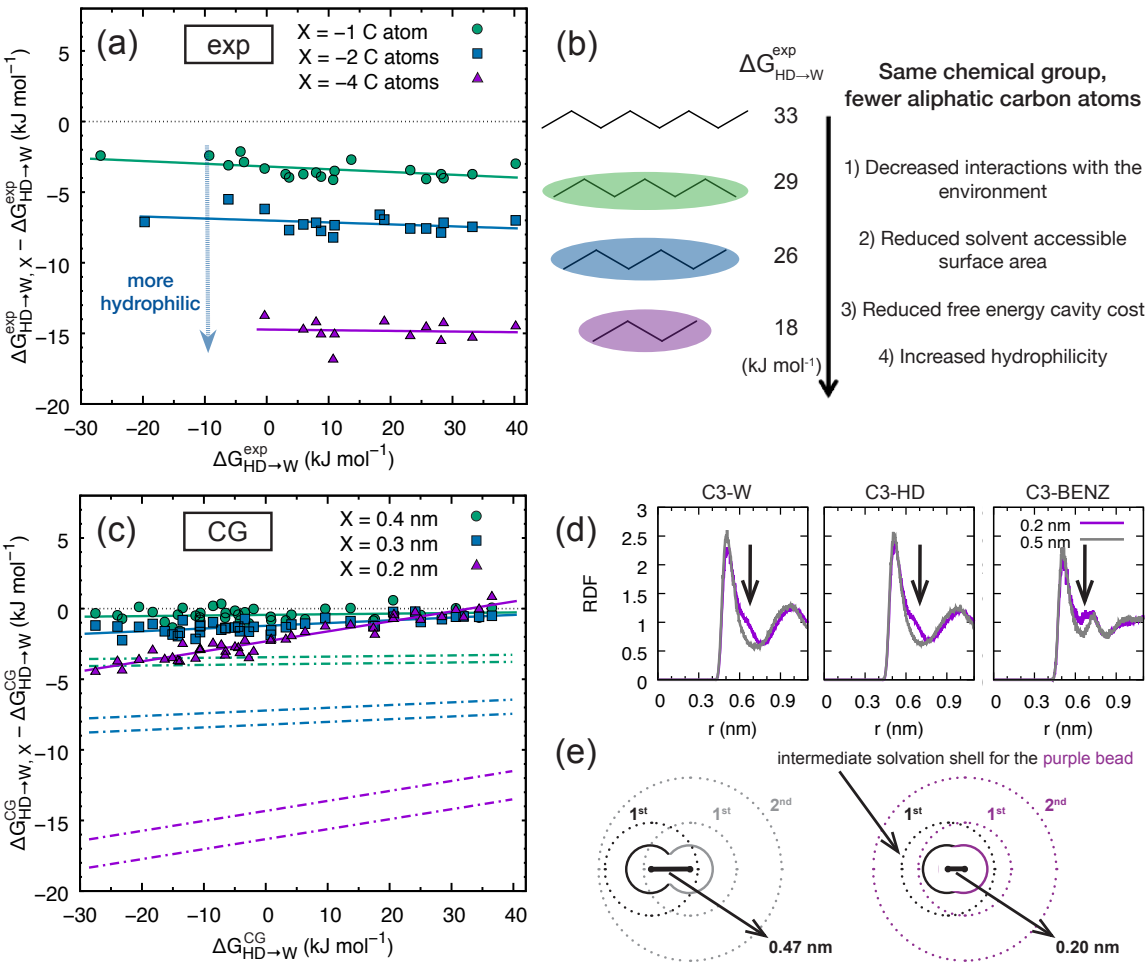


Figure 2: Experimental (top) and Martini (bottom) partitioning behavior of molecules upon removal of aliphatic carbon atoms for the same chemical functional group. (a) The hexadecane→water free energy of transfer ($\Delta G_{\mathrm{HD}\to\mathrm{W}}$) for a molecule is plotted against the difference between the same free energy and the one for the corresponding molecule (same functional group) where 1 (green circles), 2 (blue squares), and 4 (purple triangles) aliphatic carbon atoms have been removed. Fits are also shown for the various data sets. (b) Summary of how molecular properties change upon molecular size reduction, including the example of the change of $\Delta G_{\mathrm{HD}\to\mathrm{W}}$ for octane upon removal of 1, 2, and 4 carbon atoms. (c) The $\Delta G_{\mathrm{HD}\to\mathrm{W}}$ for a Martini two-bead "standard" molecule (*i.e.*, 4 atoms-to-1 CG site mapped molecules with a bond length of 0.5 nm) is plotted against the changes of the free energy of transfer upon bond length reduction to 0.4 nm (green circles), 0.3 nm (blue squares) and 0.2 nm (purple triangles), corresponding to the removal of 1, 2, and 4 aliphatic carbon atoms (that is, to 3.5-to-1, 3-to-1, and 2-to-1 atoms-to-CG sites mapping schemes, respectively). Fits are also shown for the various data sets (solid lines). The dashed lines correspond to shifts of the fits which account for Martini CG particle type changes upon size reduction (see text). (d) Solute-solvent radial distribution functions (RDFs) for a Martini two-bead molecule (in this specific example using CG particles of C3 type, but other bead types lead to the same result) solvated in water, hexadecane and benzene (from left to right) having two different bond lengths: 0.50 nm (gray) and 0.20 nm (purple). (e) Schematic of how too short a bond length affects the solvation shells and thus the interaction between the solute molecule and its environment.

the effect is *not constant* across the whole hydrophilicity scale spanned by the horizontal axis, a trend most evident when looking at the 0.2 nm case. In particular, short bond length hydrophilic molecules (left-hand side of Figure 2c) are somewhat less hydrophilic than they should (compare to Figure 2a), while hydrophobic molecules (right-hand side of Figure 2c) eventually get even more hydrophobic than their corresponding "full-size" molecule, the latter case being in *qualitative* disagreement with the experimental behavior. The same effect is qualitatively observed in all the other pairs of solvents taken into account in the original Martini parametrization,[7] as shown in Figure S2 in the Supporting Information, with variations which correlate with the relative polarity of the two solvents. We will first rationalize this effect, and then come back to the comparison with the experimental data.

The "Bond Length Effect": Increased Solute-Solvent Interactions. This behavior can be rationalized by analyzing the pair correlation functions between solute and solvent molecules and comparing the standard and short bond length cases. This is done in Figure 2d, where radial distribution functions (RDFs) are shown for a Martini two-bead molecule solvated in three different solvents (hexadecane, water, and benzene). In all cases, an extra solvation shell appears as the molecule reduces in size (see also the schematic representation shown in Figure 2e). As the peak appears, each particle of the solute effectively sees more solvent molecules, and the overall solute-solvent interaction therefore increases. This contradicts also conclusion one (Figure 2b) drawn from the experimental data set. Because different beads interact with the various solvents with different interaction strengths, this results in the imbalance across the horizontal axis observed in the 0.2 nm case in Figure 2c. The qualitatively wrong behavior of short bond length hydrophobic molecules is exactly a consequence of this: upon shrinking, a very hydrophobic solute molecule interacts more strongly with *both* water and hexadecane, but, due to a stronger interaction with hexadecane, the resulting free energy of transfer is even more hydrophobic than the "full-size" molecule.

This effect, shown for the simplest case (two-bead systems, *i.e.*, one bond length systems),

is obviously present also in systems containing more than two beads. In particular, we examined how the effect scales with the number of beads and with the geometry by computing the behavior of three-bead molecules both in a linear and ring geometry. The results are shown in Figure S3 in the Supporting Information. Notably, the effect is stronger in ring geometries than in linear ones. This can be intuitively explained in terms of the larger overlap between beads present in the case of the ring geometry, leading to a higher density of interaction sites for the same number of particles.

Implications: Short Bond Lengths Make Martini Less Intuitive. The direct comparison of Figure 2a and Figure 2c clearly show a systematic underestimation of the hydrophilicity upon molecular size reduction. However, a standard procedure in the Martini framework is to switch to a more hydrophilic particle type whenever the number of carbon atoms is reduced. For example, butane is described by a C1 particle, while propane is described by a more hydrophilic C2 bead.[7] Taking this into account, the obtained trends represented by the solid lines in Figure 2c shift to the ones depicted by the *pair of dashed lines* (assuming an average change in $\Delta G_{HD \rightarrow W}$ for a change of one bead type of 3-3.5 kJ/mol,[7] shifts of 3 and 3.5 kJ mol^{-1}, 6 and 7 kJ mol^{-1}, and 12 and 14 kJ mol^{-1} have been applied to the 0.4 nm, 0.3 nm and 0.2 nm case, respectively). This illustrates once more that the "bond length effect" has a tangible impact on the final $\Delta G_{HD \rightarrow W}$. In particular, it makes Martini less intuitive, as the choice of changing the bead type to a more hydrophilic one upon size reduction starts to depend on whether the molecule is hydrophilic (for which the "bond length effect" already accounts for some added hydrophilicity, and for which a change in bead type may be too much) or hydrophobic (for which the "bond length effect" accounts for some added hydrophobicity, and for which a change in bead type may be too little). This is relevant because in several cases short bond lengths occur when connecting two big fragments, for whose resultant macromolecule no experimental free energy is available—for example, side chain models attached to the backbone of a protein. In this case, a "naive" (that is, "just attach it and it works") building block approach would fail, as the particle type cannot be

chosen merely on the basis of the chemical group which needs to be represented. The hidden effect on partitioning behavior of shorter bond lengths blurs this intuitive building-block procedure.

3.3 Solvents with Short Bond Lengths: Excessive Cavity Cost

Having established that the use of bond distances much shorter than the ones which were used during the CG force field parametrization can have considerable effects on the parametrized behavior of a solute molecule (*i.e.*, its partitioning behavior in the case of the Martini model), we now investigate the effect of a *solvent phase* constituted by molecules with short bond lengths. We will thus see how short bond lengths affect not only the solute-solvent but also the solvent-solvent term of Eq. 3. The $\Delta G_{\mathrm{solvent-solvent}}$ is directly linked to the *free energy cost of creating a cavity* ($\Delta G_{\mathrm{cavity}}$) in the solvent: the higher the interactions between the solvent molecules, the harder it is for them to make some room for accommodating a solute molecule. We selected benzene as a representative solvent phase described with a model containing short bond lengths, both for the (albeit limited) availability of experimental partitioning data and for its importance as prototypical aromatic molecule.

Benzene/Water Partitioning: Discrepancies Between Martini and Experiments. Similarly to the previous section, we investigate what happens when computing the free energy of transfer of a two-bead system as a function of the bond length, but this time for the benzene→water (BENZ→W) case. The results, presented in the same format of Figure 2c, are shown in Figure 3b. It is evident how smaller solutes are more hydrophobic, *i.e.*, shortening bond lengths favors partitioning to the benzene phase. The question is now whether this behavior corresponds to what is observed experimentally. Figure 3a shows how experimental BENZ→W free energies of transfer change when shortening alkyl chains by 1, 2, and 4 aliphatic carbon atoms for the same chemical functional group. Due to limited availability of experimental data, we complemented the data with COSMO-RS[38] predicted free energies of transfer. More compact molecules are found to be more hydrophilic in a

similar way to what was found in the case of HD→W free energies (Figure 2a). Given that Figure 3b shows the opposite trend, we can conclude that the behavior of Martini BENZ→W free energies of transfer upon reduction of the size of the solute molecule does not agree with experimental observations.

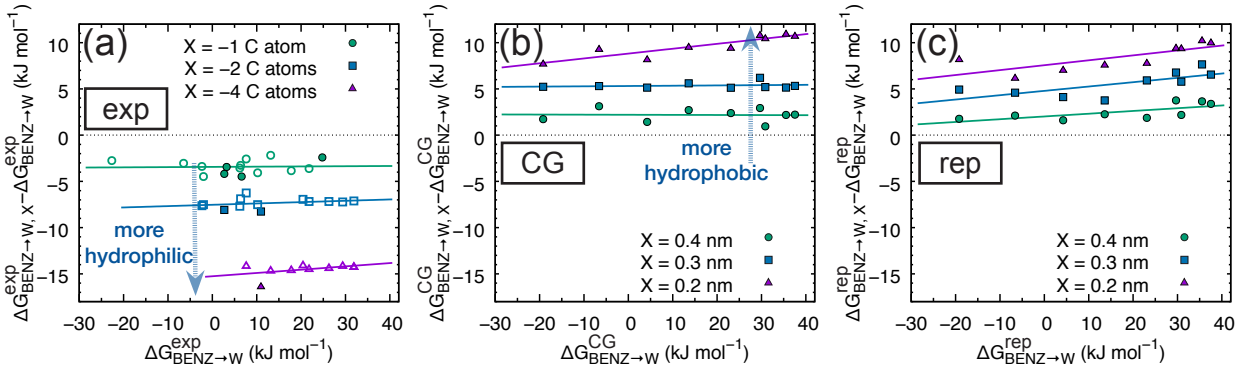


Figure 3: Experimental and Martini partitioning behavior of molecules upon removal of aliphatic carbon atoms for the same chemical functional group: short bond length phase (benzene) to water case. (a) The benzene→water free energy of transfer for a molecule is plotted against the difference between the same free energy and the one for the corresponding molecule (same functional group) where 1 (green circles), 2 (blue squares) and 4 (purple triangles) aliphatic carbon atoms have been removed. Fits are also shown for the various data sets. Experimental data points (filled) are complemented with COSMO-RS predicted (empty) free energies from Ref. 38. (b) The benzene→water free energy of transfer for a Martini two-bead "standard" molecule (*i.e.*, 4 atoms-to-1 CG site mapped molecules with a bond length of 0.5 nm) is plotted against the changes of the free energy of transfer upon bond length reduction to 0.4 nm (green circles), 0.3 nm (blue squares) and 0.2 nm (purple triangles), corresponding to the removal of 1, 2, and 4 aliphatic carbon atoms. Fits are also shown for the various data sets (solid lines). Shifts of the fits which account for Martini CG particle type changes upon size reduction (as applied in Figure 2c) are shown in Figure S4a in the Supporting Information. (c) Same plot as (b) but considering only the *repulsive* component of the LJ interaction between solvent and solute (*i.e.*, only the LJ repulsive constant C_{12} is nonzero, see also the Methods section); this is approximated as the cost of creating a cavity in the solvent.

Excessive Cavity Cost in Short Bond Length Solvents. In this section we rationalize the main cause of the discrepancies in partitioning data which involve a solvent phase with short bond lengths. With respect to the HD→W case of the previous section, we now have two different ΔG contributions which may be affected by short bond lengths at the same time: the solute-solvent and solvent-solvent terms of Eq. 3. To disentangle these two

aspects, we first performed free energy calculations treating the solute beads as purely *repulsive* particles (see also Methods). We assume these calculations to capture the free energy cost of creating a cavity in the solvents. This contribution is a proxy to the $\Delta G_{\mathrm{solvent-solvent}}$ term, as an increase in solvent-solvent interactions translates into a higher $\Delta G_{\mathrm{cavity}}$ in that solvent. The results are shown in Figure 3c. Comparison to Figure 3b points at a dominant role played by the cavity cost contribution. For a complete picture, Figure S4c (Supporting Information) shows that the *attractive* contribution (due to solute-solvent interactions, computed as the difference between the total and the repulsive one) significantly contributes only at the extreme bond length of 0.2 nm.

The major role of the cavity cost means that the prominent issue must lie with the solvent-solvent interactions. It is insightful to compare computed and experimental[39] enthalpies of vaporization ($\Delta \mathrm{H}_{\mathrm{vap}}$) for various solvents in order to determine whether solvent-solvent interactions deviate from the Martini trend in the case of short bond length phases. This is done in Figure 4a, where data points corresponding to various molecular classes have been depicted with different point symbols. We remark that enthalpies of vaporization are not parameterization targets of the Martini force field, and are systematically underestimated due to the limited fluid range of the employed 12-6 LJ potential form.[7,8] From Figure 4a, it is evident that while most solvents (such as water and the alkanes) follow the Martini trend, short bond length containing models—used to describe ring structures such as benzene within the Martini CG model—deviate from the trend and possess *systematically higher* $\Delta \mathrm{H}_{\mathrm{vap}}$. This confirms the issue with solvent-solvent interactions, which cause interactions with short bond length containing solutes to become overestimated. A few other trends are evident, as highlighted in Figure S5 and associated discussion (Supporting Information).

Both the repulsive contribution (Figure 3c) and the comparison of experimental and Martini $\Delta \mathrm{H}_{\mathrm{vap}}$ (Figure 4a) point towards the discrepancy observed in Figure 3b be rooted in the $\Delta G_{\mathrm{solvent-solvent}}$. In particular, Figure 4a showed that solvent-solvent interactions are too high in the case of a short bond length solvent phase. The consequence of this point,

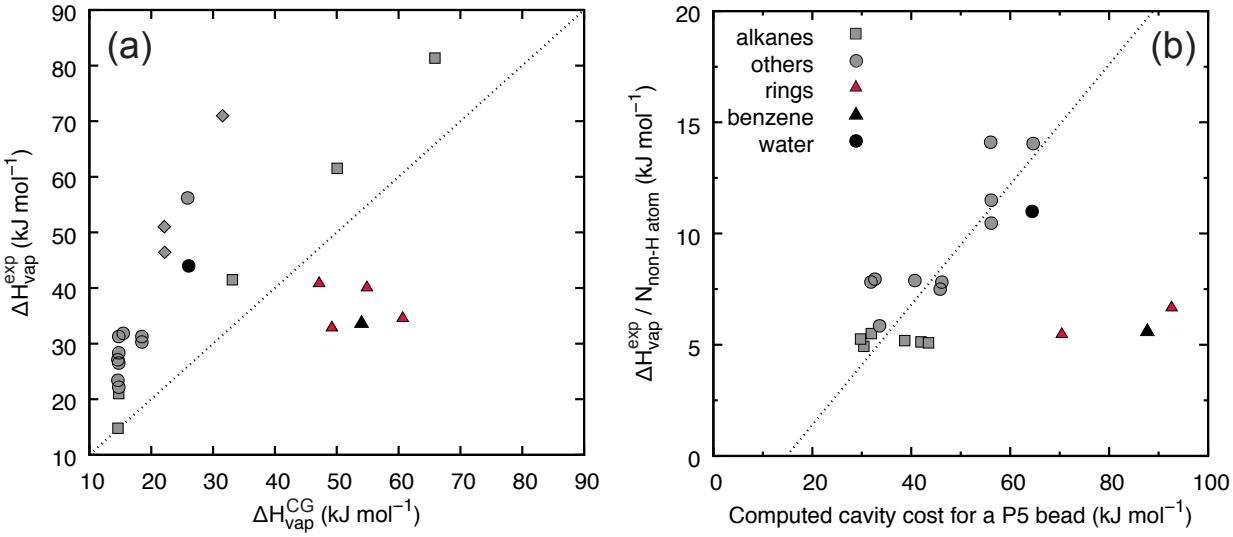


Figure 4: Behavior of enthalpies of vaporization and the cost of creating a cavity in a solvent in the Martini model. (a) Comparison of computed and experimental enthalpies of vaporization. (b) Computed free energy cost of creating a cavity (see Methods) for a P5 Martini particle type in a Martini solvent *vs.* the experimental enthalpy of vaporization of the solvent divided by the number of non-hydrogen atoms present in the molecule. In both figures, rings (described by short bond length models) do not follow the Martini trend. Benzene and water are highlighted.

as anticipated earlier, is a too high free energy cost of creating a cavity in the short bond length phase. Experimental data for such free energies are not available, while they can be approximately computed (see Methods). However, we expect them to follow trends within Martini, and to correlate with the ΔH_{vap} data. Indeed, strong interactions between solvent molecules mean both a high cost of creating a cavity in that solvent *and* a large ΔH_{vap} of that solvent. We indeed find a strong correlation between experimental ΔH_{vap} data (divided by the number of non-hydrogen atoms) and computed ΔG_{cavity} for the corresponding Martini solvent models (Figure 4b). Note that the cost of placing a bead of type P5 is plotted in this case, but results are qualitatively the same for any Martini bead type (see, for example, Figure S5b in the Supporting Information). Short bond length models indeed deviate from the trend and present considerably higher ΔG_{cavity}. In particular, we note that the *cost of creating a cavity in benzene is higher than in water* (Figure 4b, black data points). Because of this, when a solute molecule size is reduced, partitioning to the benzene phase is favored

as *one gets a bigger discount* on the cost of creating a cavity in benzene than in water. The $\Delta G_{\text{solute}-\text{solvent}}$ contribution is significant only for the extreme bond length of 0.2 nm (see Figure S4c in the Supporting Information), and it is responsible for the slope observed in the 0.2 nm data in Figure 3b. In conclusion, the main reason for the qualitatively wrong behavior of benzene→water partitioning upon shrinking of a solute molecule is the too high cost of creating a cavity in the short bond length solvent benzene.

3.4 Short Bond Lengths Caused by Weak Force Constants

We have seen how short bond lengths impact the parametrized behavior of the Martini model. Such short bond lengths can not only be achieved by setting the equilibrium bond distance to a lower value but also by weakening the force constant. In this section, we explore the effect of the force constant on the behavior of the model.

Weak Force Constants Impact the Behavior of the Martini Model. To investigate the effect of the force constant we discuss three systems of increasing complexity. The first one is a 1:1 mixture of two different three-bead molecules, namely dodecane (DOD) and dodeca-1,3,5,7-tetraene (DODE). In the Martini framework they are represented by a C1-C1-C1 and C4-C4-C1 model, respectively. We decrease the force constant of both bonds, C4-C4 and C4-C1, of the DODE model from 1250 kJ mol^{-1} (the standard Martini force constant for aliphatic chains) to 200 kJ mol^{-1}. The system, initially mixed, stays so for force constants above 500 kJ mol^{-1}. However, it starts to demix for force constants weaker than 500 kJ mol^{-1}, as can be seen by the steep decrease in number of DOD-DODE contacts reported in Figure 5a as a function of the force constant. Lowering the force constant for the bonds in one of the molecules is thus found to induce phase separation.

For the second test case, we constructed an icosahedron of P4 beads including a central P4 bead and solvated eight of them in water—which is also described by P4 beads in Martini. All twelve outer beads of each icosahedron are connected to the central bead by a harmonic potential with a force constant of 1250 kJ mol^{-1} and a minimum position at 0.47 nm. In

addition, six bonds exist on the surface connecting pairs of adjacent corners (red bonds in the inset of Figure 5b). Thus, each corner is connected to exactly one neighboring corner. In our simulations we varied the force constant of the surface bonds while keeping the force constants to the center constant. This ensures that the overall size of the icosahedron stays approximately unaffected while effective bond lengths on the surface can impact the interaction with the environment. We simulated the association of the eight icosahedrons in water (simulation time 500 ns) with different force constant of the surface bonds. Figure 5b shows the normalized cluster size of the eight icosahedrons for the different simulations. When applying a force constant of 563 kJ mol^{-1} at the icosahedron surface, the cluster size distribution changes. Our simulations show that short bond length regions on the surface of a given icosahedron interact particularly with such short bond length regions of other icosahedrons. As a result, cluster formation is strongly enhanced.

The third system consists of nine polyleucine transmembrane α-helical peptides embedded in a DPPC bilayer (starting as monomers, see Figure S6b in the Supporting Information). We compare a protein model with an elastic network in the protein backbone using a common force constant of 500 kJ mol^{-1} and a protein model without elastic network. Again, the number of protein clusters is analyzed to investigate if the elastic network impacts the protein-protein interaction. Figure 5d depicts the number of clusters present during a 10 μs simulation. Evidently, the system simulated without elastic network (black line) consists of more and smaller clusters while in the system with the elastic network (red line) only two clusters are present after 10 μs. Even more concerning is the fact that the system is composed of one monomer and one octamer. Comparison to experimental data suggests that the protein model without elastic network is already too sticky. It slightly underestimates the population of the monomeric state.[40–42] For completeness, we also performed simulations of polyleucine with the ElNeDyn[34] model and a standard force constant of 500 kJ mol^{-1}. The number of clusters being present during the 10 μs simulation are shown in Figure S6c. The cluster formation happens slightly faster than in the case of Martini 2.2 with elastic network

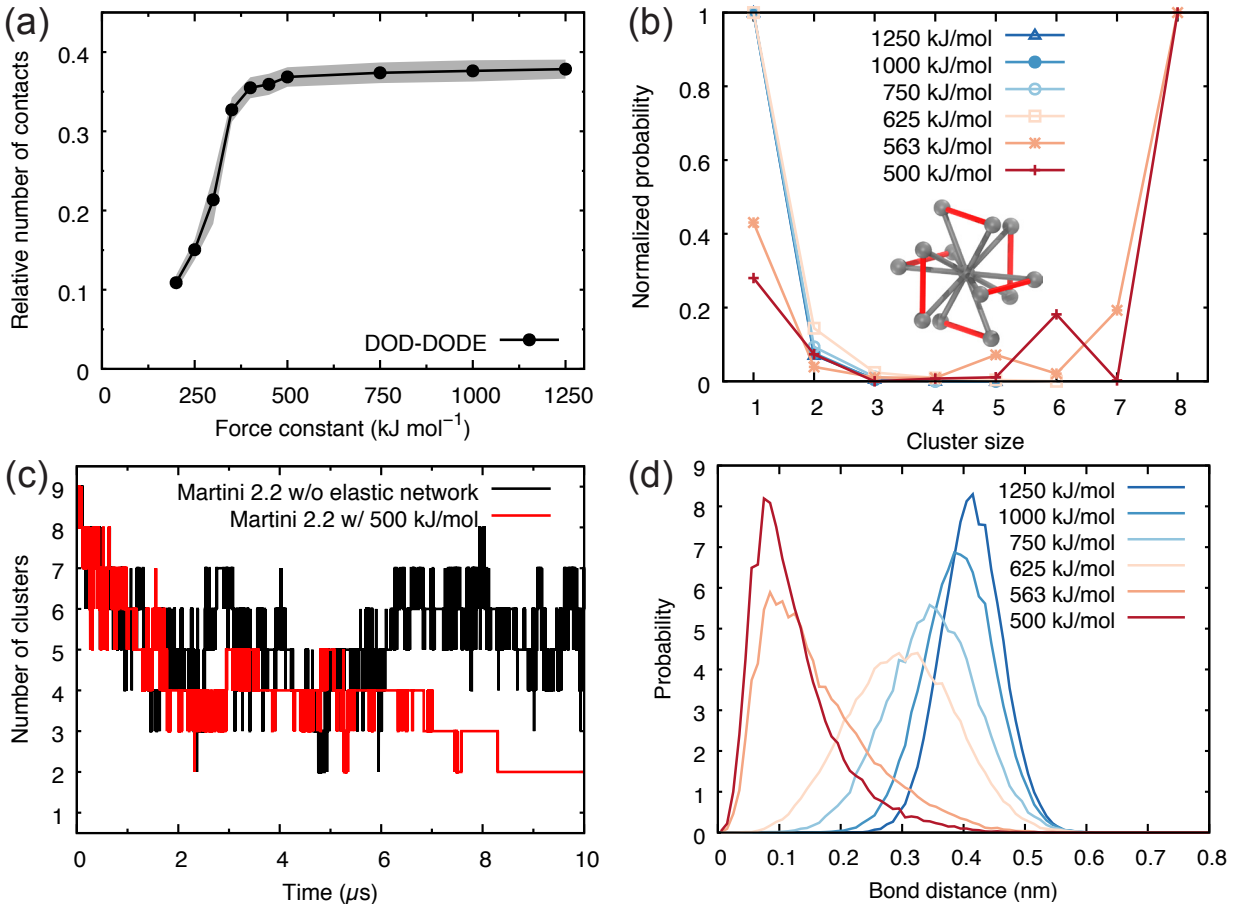


Figure 5: Effect of weak bond force constants on the behavior of Martini systems of increasing complexity. (a) Relative number of DOD-DODE contacts (number of DOD-DODE contacts over the total number of contacts made by DODE molecules) in a 1:1 DOD:DODE mixture as a function of the force constant used in the two bonds of three-bead DODE molecule. The corresponding effective bond length distributions for such bonds are shown in Figure S6a. (b) Cluster size distribution for a simulation of eight icosahedrons (inset) described by P4 Martini beads in water (also described by P4 beads) as a function of the force constant used for the six bonds on the surface of the icosahedrons (red bonds in the inset)—and (d) corresponding effective bond length distributions for such bonds. (c) Number of clusters being present in a simulation of nine polyleucine transmembrane α-helical peptides embedded in a DPPC bilayer modeled without (black line) and with (red line) an elastic network using a common force constant of 500 kJ mol^{-1} (see also Methods).

and the clusters appear to be more stable.

Weak Force Constants Lead to the Formation of Super-Interaction Centers. The behavior observed for these three systems, *i.e.*, increased aggregation between models which use weaker force constants, can be rationalized in terms of the effective bond length distributions present in the systems. Such distributions are shown in Figure 5d for the icosahedron case, but they are qualitatively the same for the two other systems (see also Figure S6a in the Supporting Information). The equilibrium bond distance of the harmonic bond being fixed to 0.47 nm for all the bonds of Figure 5d, the effective bond length in the systems differ considerably when force constants lower than 1000 kJ mol^{-1} are used. The weaker the force constant used, the shorter the effective bond length.

Weak force constants let the CG beads within a molecule get too close; when that happens, their LJ interactions with a third bead in the surrounding add up. This increased interaction with the environment results in the creation of a super-interaction center, that is a region with high density of interaction sites—analogous to the situation described in Section 3.2 and 3.3. However, for the creation of such a super-interaction center not only the equilibrium position of the bonded potential (as seen in Section 3.2 and 3.3) is of importance but also its force constant. If this force constant is too weak, the bonded interaction cannot compete with the non-bonded force. Their imbalance enables the bonded beads to approach closely resulting in a distance distribution which is not centered at the minimum position anymore (Figure 5d and S6a). Thus, there is a lower limit for the force constant of bonded interactions described by harmonic potentials if they entail exclusions, *i.e.*, non-bonded interactions between the bonded beads are not present. If the lower limit is undercut, the harmonic potential cannot compete with the non-bonded potentials which thus give rise to collapse of the beads connected by the weak force constant potential.

4 Outlook

We have seen how the lack of size-dependent Lennard-Jones parameters can artificially increase the barrier in free energy profiles of dimerization. This effect is larger, the larger the mismatch between the solute-solute and solute-solvent Lennard-Jones sigma parameters, and the bigger the solute molecules. We have then investigated in detail the effect that the use of bond lengths shorter than the ones used during the parametrization has on the behavior of the Martini CG force field. Shortening bond lengths increases the density of CG interaction sites and may thus lead to imbalances. In particular, we have seen that shortening the bond length of a solute molecule increases its interactions with any solvent. Because different beads interact with the various solvents with different interaction strengths, the effect is nonlinear, and thus can unbalance the carefully parametrized behavior of the Martini force field. We have also shown how the use of a solvent phase constituted of short bond length molecules leads to even bigger discrepancies. The enhanced interactions between solvent molecules increases the cost of creating a cavity in the short bond length phase disproportionately, leading again to an unbalanced behavior of such phases. Finally, we have seen how there is a lower limit for the force constant of bonded interactions described by harmonic potentials if they entail exclusions, *i.e.*, non-bonded interactions between the bonded beads are not present. If the lower limit is undercut, the harmonic potential cannot compete with the non-bonded potentials which leads to short bond lengths and thus increased interactions.

Implications for the Current Use of Martini 2. Discrepancies between parametrized and observed behavior may arise when systems are rich in molecules containing short bond lengths. A short bond length phase clearly possesses increased interactions both with eventual solute molecules and between the molecules of the phase themselves. Such short bond lengths arise in Martini models when finer mappings are designed. A perfect match with an atomistic bond distribution should be sacrificed in exchange for more reasonable densities of CG interaction sites. Deviations from the parametrized behavior are mostly expected when mixing standard and short bond length systems. A consistent use of short bond lengths for

both solute(s) and solvent(s) may reduce the discrepancies observed in properties such as partitioning or mixing due to consistent shifts in overall behavior. However, properties of models rich in short bond lengths may deviate from the overall behavior of the force field.[12,43–45] Moreover, as soon as there are standard and short bond length models together, short bond length molecules will interact predominantly with other short bond length molecules. This effect may be partly responsible for the need of "custom" beads emerged when modeling a number of polymers. Such systems rely heavily on S-beads, hence contain short bond lengths, and need to behave properly in both S-beads and regular solvents.[46,47] This effect may also contribute to the observed stickiness of Martini proteins[48–50] or sugars.[51] While this complex multicomponent problem is not straightforward and affects also atomistic force fields,[52] short bond lengths will be part of the problem in the case of Martini, as both sugars and proteins contain short bonds.

More generally, when dealing with models based on a building block approach, not only the calibration of the fragments but also their connection must be considered carefully. Despite careful calibration of the fragments, their connection can introduce artefacts, as it was shown to be the case in the Martini model. The extensive use of a certain model within a wide and various research community can only be beneficial to the improvement of the model, as such nontrivial effects can be spotted more promptly. In a broader view, this can affect also atomistic force fields based on similar building block philosophies. While there is much less variability between bond lengths at atomistic resolution, a similar role to the one played by bonds within Martini may be played by dihedral angles in atomistic force fields.

Directions for Reparametrizations. The findings reported in this work lead to clear paths for improvements of the Martini CG model, and should be also taken into account in the parametrization of any other building block based force field. Specifically, size-dependent Lennard-Jones parameters are necessary to ensure balanced interactions between CG interaction sites of different sizes and to avoid artifacts such as increased barriers in dimerization profiles. The density of interaction sites is a very critical property of the system. If finer

mappings are required due to symmetry or necessity of a description with a higher resolution, well calibrated particles with different sizes should be available. Such bead sizes should probably be calibrated in a way that will lead to correct trends for enthalpies of vaporization (and hence cavity costs) for different resolutions. Ideally, models of the same molecules with different resolutions, *e.g.*, a dodecane molecule mapped with three 4-to-1, or four 3-to-1, or six 2-to-1 atoms-to-CG-site should give the same enthalpy of vaporization, free energy of solvation (hence cavity cost) and hence mix ideally between themselves. The different resolutions are intristically coupled to the bond lengths used in the systems. If short bond lengths are necessary, it is because finer mappings or very branched chemical moieties are being represented. Thus, finer mappings imply smaller beads and shorter bond lengths, while coarser ones imply larger beads and longer bond lengths. If this harmony is not maintained, an imbalance in the parametrized behavior of the model is expected. Lastly, the elastic network approach might be replaced by a Go-model approach[53] to (i) avoid prolbems with weak bonds, and (ii) allow some folding-unfolding at the same time.

These guidelines have been taken into account in the reparametrization of the Martini CG force field which led to the very recent development of Martini 3.0.[54]

Acknowledgement

R.A. thanks The Netherlands Organisation for Scientific Research NWO (Graduate Programme Advanced Materials, No. 022.005.006) for financial support, Jaakko J. Uusitalo for pioneering insights on the misbehavior of Martini S-bead systems, and Ignacio Faustino for critical reading of the manuscript. S.T. thanks the European Commission for financial support via a Marie Sklodowka-Curie Actions Individual Fellowship (MicroMod-PSII, grant agreement 748895).

Supporting Information Available

PMFs of dimerization for single beads in water, and for pyrene in water; partitioning behavior of two-bead CG models upon bond length reduction between other solvents (water/chloroform, water/octanol, and water/ether); hexadecane/water partitioning behavior of linear and cyclic three-bead CG models upon bond length reduction; contributions to the CG benzene/water free energy of transfer change upon bond length reduction; further details on the comparison between experimental and Martini enthalpies of vaporization; free energy cavity cost for a Martini C1 bead; effective bond length distributions for the DOD:DODE mixture; rendering of the starting configuration of the polyleucine system.

This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Klein, M. L.; Shinoda, W. *Science* **2008**, *321*, 798–800.

(2) Voth, G. A. *Coarse-graining of condensed phase and biomolecular systems*; CRC press, 2008.

(3) Noid, W. G. *J. Chem. Phys.* **2013**, *139*, 090901.

(4) Brini, E.; Algaer, E. A.; Ganguly, P.; Li, C.; Rodríguez-Ropero, F.; van der Vegt, N. F. A. *Soft Matter* **2013**, *9*, 2108–2119.

(5) Ingólfsson, H. I.; López, C. A.; Uusitalo, J. J.; de Jong, D. H.; Gopal, S. M.; Periole, X.; Marrink, S. J. *WIREs Comput. Mol. Sci.* **2014**, *4*, 225–248.

(6) Pak, A. J.; Voth, G. A. *Curr. Opin. Struct. Biol.* **2018**, *52*, 119–126.

(7) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.

(8) Marrink, S. J.; Tieleman, D. P. *Chem. Soc. Rev.* **2013**, *42*, 6801–6822.

(9) Marrink, S. J.; de Vries, A. H.; Mark, A. E. *J. Phys. Chem. B* **2004**, *108*, 750–760.

(10) Vögele, M.; Holm, C.; Smiatek, J. *J. Chem. Phys.* **2015**, *143*, 243151.

(11) Winands, T.; Bockmann, M.; Schemme, T.; Ly, P.-M. T.; de Jong, D. H.; Wang, Z.; Denz, C.; Heuer, A.; Doltsinis, N. L. *Phys. Chem. Chem. Phys.* **2016**, *18*, 6217–6227.

(12) Alessandri, R.; Uusitalo, J. J.; de Vries, A. H.; Havenith, R. W. A.; Marrink, S. J. *J. Am. Chem. Soc.* **2017**, *139*, 3697–3705.

(13) Bochicchio, D.; Pavan, G. M. *ACS Nano* **2017**, *11*, 1000–1011.

(14) Frederix, P. W. J. M.; Patmanidis, I.; Marrink, S. J. *Chem. Soc. Rev.* **2018**, *47*, 3470–3489.

(15) Modarresi, M.; Franco-Gonzalez, J. F.; Zozoulenko, I. *Phys. Chem. Chem. Phys.* **2018**, *20*, 17188–17198.

(16) Uusitalo, J. J.; Ingólfsson, H. I.; Akhshi, P.; Tieleman, D. P.; Marrink, S. J. *J. Chem. Theory Comput.* **2015**, *11*, 3932–3945.

(17) de Jong, D. H.; Baoukina, S.; Ingólfsson, H. I.; Marrink, S. J. *Comput. Phys. Commun.* **2016**, *199*, 1–7.

(18) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.

(19) Malde, A. K.; Zuo, L.; Breeze, M.; Stroet, M.; Poger, D.; Nair, P. C.; Oostenbrink, C.; Mark, A. E. *J. Chem. Theory Comput.* **2011**, *7*, 4026–4037.

(20) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(21) de Jong, D. H.; Singh, G.; Bennett, W. D.; Arnarez, C.; Wassenaar, T. A.; Schafer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. *J. Chem. Theory Comput.* **2012**, *9*, 687–697.

(22) Nosé, S. *Mol. Phys.* **1984**, *52*, 255–268.

(23) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.

(24) Parrinello, M.; Rahman, A. *J. App. Phys.* **1981**, *52*, 7182–7190.

(25) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.

(26) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. *SoftwareX* **2015**, *1*, 19–25.

(27) Torrie, G.; Valleau, J. *J. Comput. Phys.* **1977**, *23*, 187–199.

(28) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Pullman, B., Ed.; Springer Netherlands: Dordrecht, 1981; pp 331–342.

(29) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(30) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

(31) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1994**, *222*, 529–539.

(32) Shirts, M. R.; Chodera, J. D. *J. Chem. Phys.* **2008**, *129*, 124105.

(33) Wassenaar, T. A.; Ingólfsson, H. I.; Böckmann, R. A.; Tieleman, D. P.; Marrink, S. J. *J. Chem. Theory Comput.* **2015**, *11*, 2144–2155.

(34) Periole, X.; Cavalli, M.; Marrink, S.-J.; Ceruso, M. A. *J. Chem. Theory Comput.* **2009**, *5*, 2531–2543.

(35) Southall, N. T.; Dill, K. A.; Haymet, A. D. J. *J. Phys. Chem. B* **2002**, *106*, 521–533.

(36) Natesan, S.; Wang, Z.; Lukacova, V.; Peng, M.; Subramaniam, R.; Lynch, S.; Balaz, S. *J. Chem. Inf. Model.* **2013**, *53*, 1424–1435.

(37) Mannhold, R.; Poda, G. I.; Ostermann, C.; Tetko, I. V. *J. Pharm. Sci.* **2009**, *98*, 861–893.

(38) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. *J. Phys. Chem. A* **1998**, *102*, 5074–5085.

(39) Haynes, W. M. *CRC handbook of chemistry and physics*; CRC press, 2014.

(40) Zhou, F. X.; Cocco, M. J.; Russ, W. P.; Brunger, A. T.; Engelman, D. M. *Nat. Struct. Mol. Biol.* **2000**, *7*, 154.

(41) Zhou, F. X.; Merianos, H. J.; Brunger, A. T.; Engelman, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2250–2255.

(42) Grau, B.; Javanainen, M.; García-Murria, M. J.; Kulig, W.; Vattulainen, I.; Mingarro, I.; Martínez-Gil, L. *Cell Stress* **2017**, *1*, 90–106.

(43) Melo, M. N.; Ingólfsson, H. I.; Marrink, S. J. *J. Chem. Phys.* **2015**, *143*, 243152.

(44) Bereau, T.; Kremer, K. *J. Chem. Theory Comput.* **2015**, *11*, 2783–2791.

(45) Genheden, S. *J. Comput. Aided Mol. Des.* **2017**, *31*, 867–876.

(46) Rossi, G.; Monticelli, L.; Puisto, S. R.; Vattulainen, I.; Ala-Nissila, T. *Soft Matter* **2011**, *7*, 698–708.

(47) Grunewald, F.; Rossi, G.; de Vries, A. H.; Marrink, S. J.; Monticelli, L. *J. Phys. Chem. B* **2018**, *122*, 7436–7449.

(48) Stark, A. C.; Andrews, C. T.; Elcock, A. H. *J. Chem. Theory Comput.* **2013**, *9*, 4176–4185.

(49) Javanainen, M.; Martinez-Seara, H.; Vattulainen, I. *PLoS One* **2017**, *12*, 1–20.

(50) Periole, X.; Zeppelin, T.; Schiøtt, B. *Sci. Rep.* **2018**, *8*, 5080.

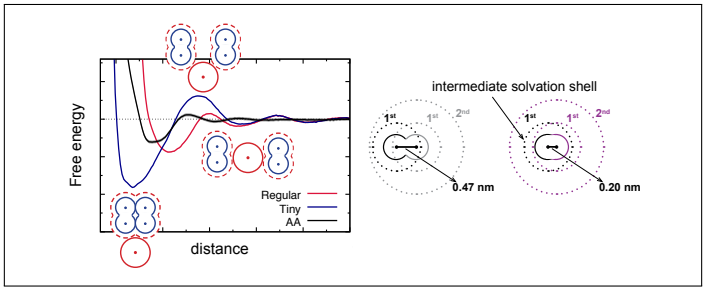
(51) Schmalhorst, P. S.; Deluweit, F.; Scherrers, R.; Heisenberg, C.-P.; Sikora, M. *J. Chem. Theory Comput.* **2017**, *13*, 5039–5053.

(52) Petrov, D.; Zagrovic, B. *PLOS Comput. Biol.* **2014**, *10*, 1–11.

(53) Poma, A. B.; Cieplak, M.; Theodorakis, P. E. *J. Chem. Theory Comput.* **2017**, *13*, 1366–1374.

(54) Souza, P. C. T.; ...,; Marrink, S. J. *in preparation*

Graphical TOC Entry



pitfalls-achemso.pdf (892.55 KiB) view on ChemRxiv • download file

Supporting information for:

Pitfalls of the Martini Model

Riccardo Alessandri,[†,‡] Paulo C. T. Souza,[†,‡] Sebastian Thallmair,[†] Manuel N. Melo,[¶] Alex H. de Vries,[†] and Siewert J. Marrink[*,†]

†*Groningen Biomolecular Sciences and Biotechnology Institute and Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands*

‡*These two authors contributed equally.*

¶*Instituto de Tecnologia Química Biológica António Xavier, Universidade Nova de Lisboa, Av. da República, 2780-157 Oeiras, Portugal*

E-mail: s.j.marrink@rug.nl

1 Differences in Bead Sizes: the Desolvation Problem

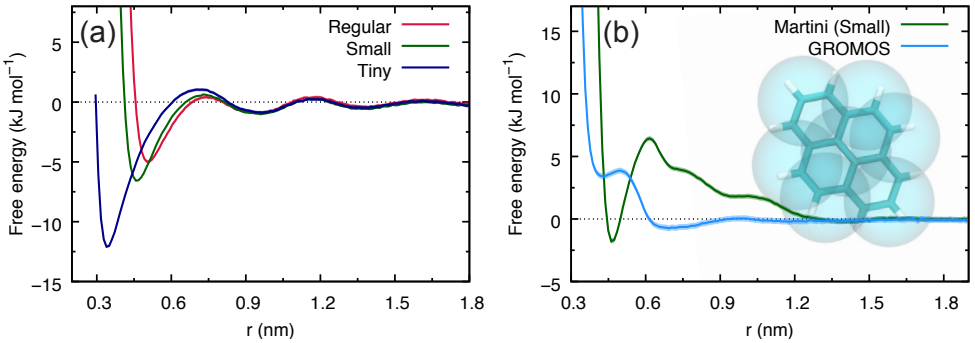


Figure S1: (a) PMF of dimerization for single beads in water. Profiles for the dimerization of two C5 beads (red), two SC5 beads (green) and two TC5 beads (blue) are shown. (b) PMF of dimerization for pyrene in chloroform. In the Martini CG model (green curve), pyrene is described by 7 SC5 beads (see inset), while the GROMOS (53A6) model (blue curve)—retrieved from the ATB server[S1]—is used for the atomistic case.

2 Solutes with Short Bond Lengths

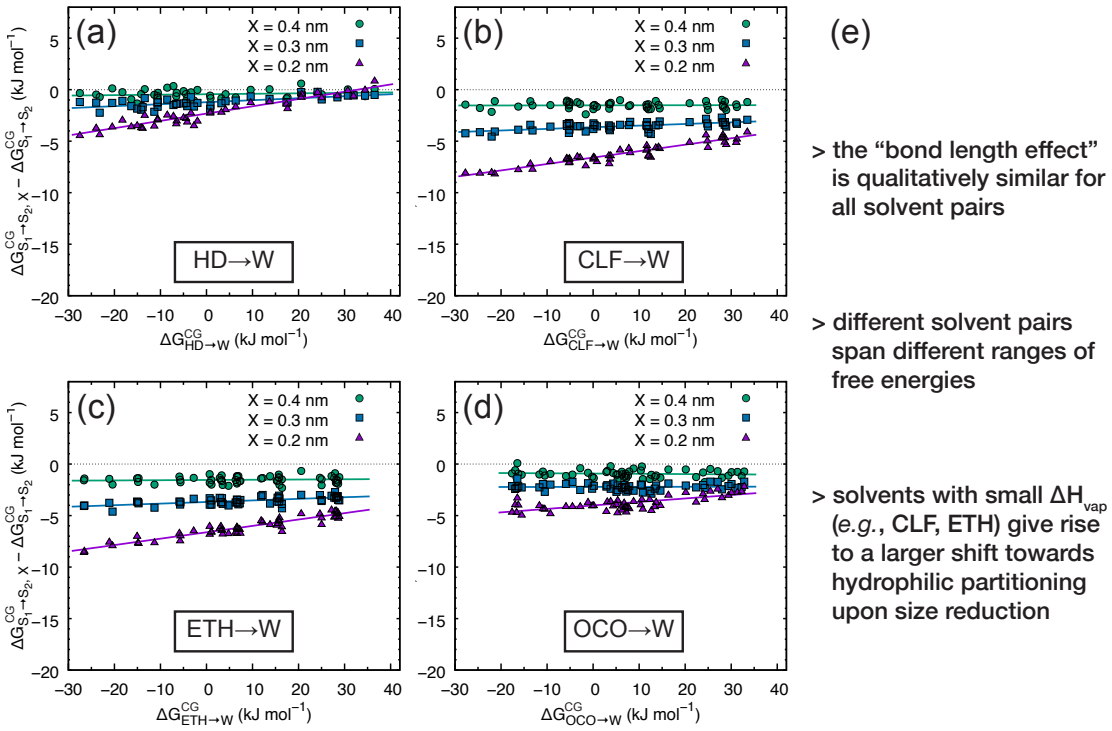


Figure S2: (a) Same as Figure 2c of the main text; (b)-(d) same plots as (a) but for chloroform (CLF), ether (ETH) and octanol (OCO) to water partitioning. (e) Trends which can be extracted from (a)-(d).

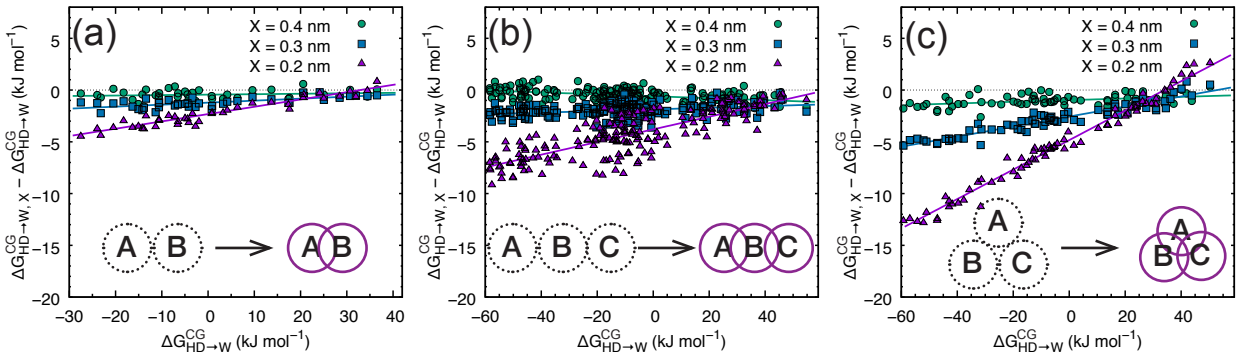


Figure S3: Scaling of the "bond length effect" with the size and geometry of Martini small molecules. (a) Same as Figure 2c of the main text. (b) Same plot as (a) but for three-bead linear models (see inset) representing linear molecules containing 12 non-hydrogen atoms. (c) Same plot as (a) but for three-bead ring models (see inset) representing 12 non-hydrogen atoms ring molecules.

3 Solvents with Short Bond Lengths

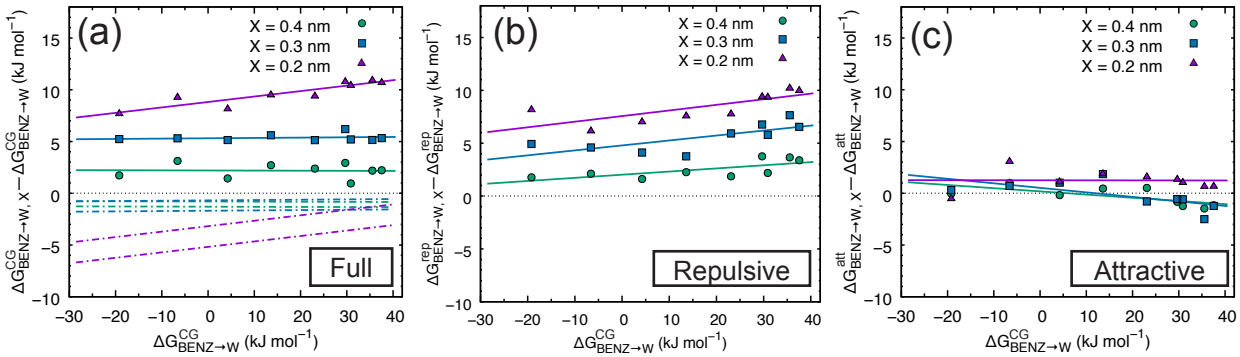


Figure S4: Contributions to the benzene→water free energies of transfer in the Martini CG model. (a) Total (same as Figure 3b of the main text, but including also the shifts as applied in Figure 2c of the main text), (b) repulsive (same as Figure 3c of the main text), and (c) attractive contributions. The latter contribution was computed as the difference between the total and the repulsive ones. The repulsive component of the Lennard-Jones, which we take as an approximate estimation of the cavity cost, causes the more and more favorable partitioning to the benzene phase as the molecular size decreases. The attractive part is mostly constant and only increases in the case of the smallest bond length.

Martini Enthalpies of Vaporization. We notice other trends from Figure S5a (which is the same as Figure 4a of the main text, but with other chemical classes highlighted). First of all, we note that molecules containing atoms belonging to the third period of the periodic table fall in the upper-left edge of the trend. This indicates a relative underestimation of the ΔH_{vap} for these molecules. We ascribe this to the presence of the heavier atoms (Cl, or S), which due to their higher polarizability and mass are likely to increase the ΔH_{vap} of the molecule with respect to the one predicted by the Martini particle used to describe it - being Martini mostly calibrated with respect to second period elements such as N, C, and O. The enthalpies of vaporization of alcohols are also underestimated, probably due to underestimated hydrogen-bonding capabilities (likely to increase ΔH_{vap}'s) of the Martini models.

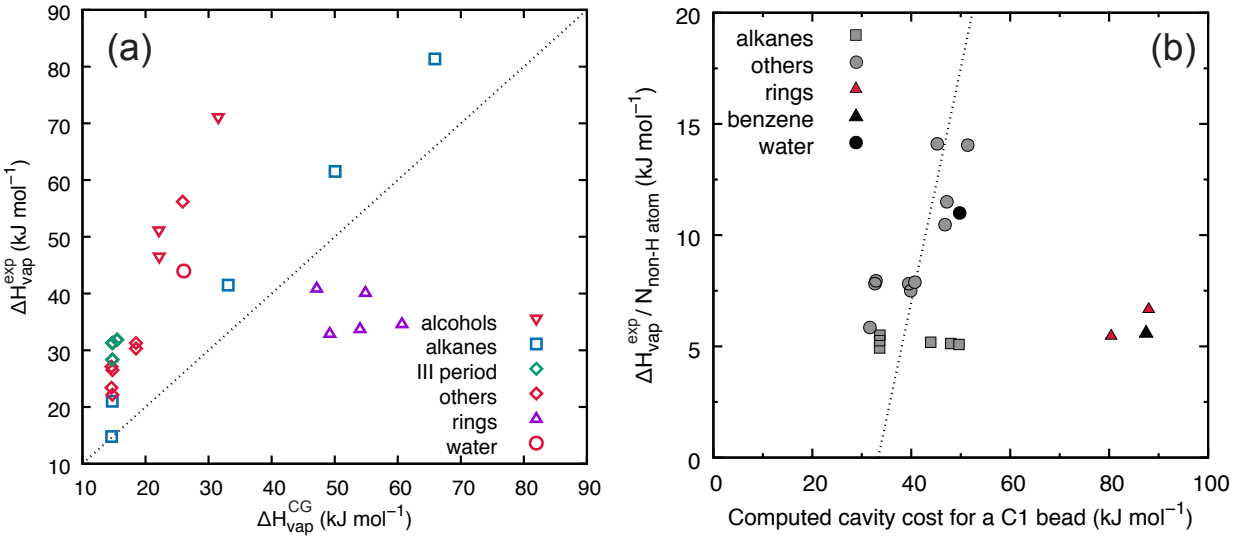


Figure S5: Behavior of enthalpies of vaporization and the cost of creating a cavity in a solvent in the Martini model. (a) Comparison of computed and experimental enthalpies of vaporization. A few chemical classes are indicated with different data points (see legend). (b) Same plot as Figure 4b of the main text but for a Martini C1 bead.

4 Short Bond Lengths Caused by Weak Force Constants

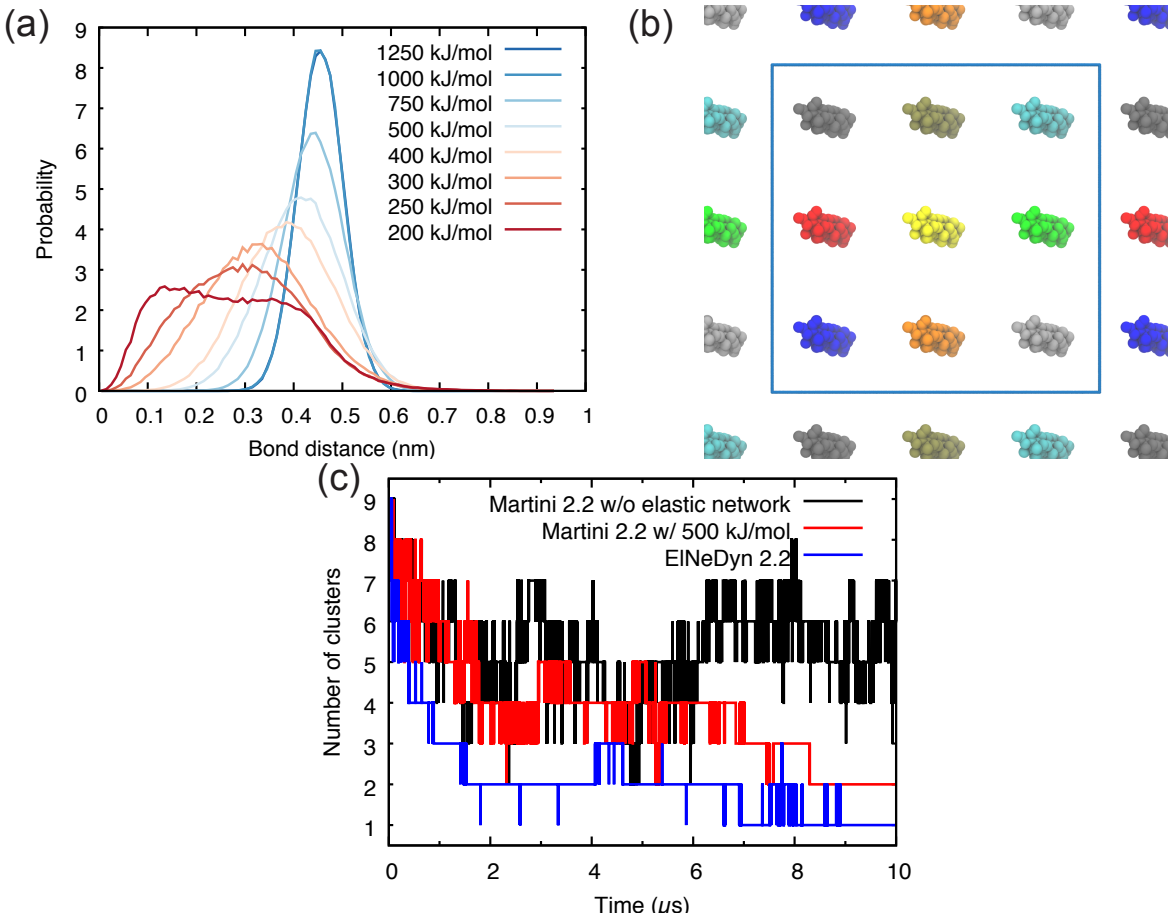


Figure S6: (a) Effective bond length distribution as a function of force constant for DODE. DODE is part of a 1:1 DOD:DODE mixture as described in section 3.4 of the main text. (b) Top view of the starting configuration for the nine polyleucines embedded in a DPPC bilayer (section 3.4 of the main text). Only the peptides are rendered.

References

(S1) Malde, A. K.; Zuo, L.; Breeze, M.; Stroet, M.; Poger, D.; Nair, P. C.; Oostenbrink, C.; Mark, A. E. *J. Chem. Theory Comput.* **2011**, *7*, 4026–4037.

pitfalls-achemso-suppinfo.pdf (9.24 MiB) view on ChemRxiv • download file